

Best methods for evaluating educational impact: a comparison of the efficacy of commonly used measures of library instruction

Katherine Schilling, MLS, EdD; Rachel Applegate, MLS, PhD

See end of article for authors' affiliations.

DOI: <http://dx.doi.org/10.3163/1536-5050.100.4.007>

Objectives and Background: Libraries are increasingly called upon to demonstrate student learning outcomes and the tangible benefits of library educational programs. This study reviewed and compared the efficacy of traditionally used measures for assessing library instruction, examining the benefits and drawbacks of assessment measures and exploring the extent to which knowledge, attitudes, and behaviors actually paralleled demonstrated skill levels.

Methods: An overview of recent literature on the evaluation of information literacy education addressed these questions: (1) What evaluation measures are commonly used for evaluating library instruction? (2) What are the pros and cons of popular evaluation measures? (3) What are the relationships

between measures of skills versus measures of attitudes and behavior? Research outcomes were used to identify relationships between measures of attitudes, behaviors, and skills, which are typically gathered via attitudinal surveys, written skills tests, or graded exercises.

Results and Conclusions: Results provide useful information about the efficacy of instructional evaluation methods, including showing significant disparities between attitudes, skills, and information usage behaviors. This information can be used by librarians to implement the most appropriate evaluation methods for measuring important variables that accurately demonstrate students' attitudes, behaviors, or skills.

INTRODUCTION

In academic and professional environments that are increasingly evidence based and outcomes driven, librarians are likewise called upon to provide tangible evidence that instruction in information literacy (IL) skills is valid and legitimate. Legitimacy is usually quantified through a hierarchy of expected educational standards and outcomes. These include the impact of instruction on students' information retrieval skills, course grades, IL skills, and achievement of program and national standards. The assessment of student learning is important for demonstrating academic achievement and program success, particularly in the context of increasing tuition costs. Measuring student learning outcomes is a process that necessitates sequential and systematic evaluation of library training, workshops, or courses.

Quality educational program evaluation includes both quantitative and qualitative measures. Many commonly used evaluation tools such as self-reported perception surveys focus solely on students' perceptions of their own skills, their knowledge, or the library. These are important areas of investigation, for user satisfaction and self-confidence are significant factors in understanding students' attitudes about libraries and information. At the same time, a balanced approach is ideal, requiring librarians to distinguish between and appropriately apply effective evaluation measures

to provide evidence of both attitudes and actual learning outcomes.

Highlights

- Comparisons of library instruction evaluation measures demonstrate what surveys, written tests, practical exercises, self-assessments, and so on actually do and do not measure.
- Students routinely overestimate their information retrieval and information literacy skills.
- Students' attitudes about the library and librarians may not correlate with their perceptions of their skills and demonstrated skills.
- Research identifies a disconnect between theoretical knowledge and demonstrated skills.

Implications

- Formative assessment is not only effective for collecting baseline information, but for giving students a realistic picture of their true skill levels.
- Longitudinal summative assessment of practical skills is the truest measure of learning.
- Libraries should implement appropriate performance, affective, and behavioral measures to provide a complete and accurate assessment of learners' information literacy skills and attitudes.
- Simple descriptive statistics are not adequate for showing how evaluation measures are interdependent, which have cause-and-effect relationships, and what combinations of measures represent skills, knowledge, and learning most accurately.



This article has been approved for the Medical Library Association's Independent Reading Program <<http://www.mlanet.org/education/irp/>>.

Table 1
Evaluation measures frequently represented in the literature: 2007–2012

Author, publication date	Review	Practical exercises	Written knowledge tests	Self-reported skills	Portfolio, products, journals, logs	Citation analysis	Attitude surveys	Library usage
Appelt and Pendell, 2010 [58]				X			X	
Ariew and Lener, 2007 [59]							X	
Barratt, Nielsen, Desmet, and Balthazor, 2009 [41]					X	X		
Blummer, 2007 [45]								X
Burkhardt, 2007 [9]		X	X		X			
Cameron, Wise, and Lottridge, 2007 [8]		X						
Canick, 2009 [47]		X						
Chen, Chan, and Kong, 2010 [42]			X		X			
Crawford, 2007 [61]	X							
Figa, Bone, and MacPherson, 2009 [44]							X	
Gilstrap and Dupre, 2008 [40]					X		X	
Hernon, 2009 [46]	X							
Hufford, 2010 [37]				X				
Kisby, 2011 [64]								X
Lym, Grossman, Yannotta, and Tali, 2009 [7]		X						
Oakleaf, 2008 [33]	X							
Oakleaf, 2011 [34]	X							
Polkinghorne and Wilton, 2010 [11]					X			
Scharf, Elliot, Huey, Briller, and Joshi, 2007 [10]						X		
Schilling and Applegate, 2007 [57]		X	X	X		X	X	X
Sharma, 2007 [12]			X		X	X		
Sobel and Wolf, 2011 [56]			X	X	X			X
Staley, Branch, and Hewitt, 2010 [39]			X					
Weisskirch and Silveria, 2007 [60]		X			X		X	
Xin and Kwangsu, 2010 [65]					X			
Zhang, Watson, and Banfield, 2007 [35]	X							
Zoellner, Samson, and Hines, 2008 [38]			X	X			X	X

The authors surveyed the academic library literature on student learning assessment from 2007 to 2012 to uncover meaningful relationships between various measures of learning for assessing knowledge, skills, attitudes, or behaviors. Several questions related to the evaluation of library instruction were asked:

1. What evaluation measures are commonly used?
2. What are the pros and cons of popular evaluation measures?
3. What are the relationships between measures of skills *versus* measures of attitudes and behavior?

This review includes the articles that specifically address the use of evaluation measurements and instruments to assess IL skills or the impact of IL training on student learning. The review is not limited to the biomedical journal literature, but also includes literature from the arts, humanities, and social sciences. The resulting sample includes articles published since 2007 that report on the use of measures of performance (tests, products, portfolios), attitude (surveys), and behavior (in relation to the use of library resources). These measures are critiqued based on the extent to which they uncover meaningful relationships about learners' skills, knowledge, attitudes, and behaviors. An overview of their pros and cons, benefits, and drawbacks is given. This paper also highlights the literature that makes the best cases for drawing valid conclusions about the tangible impact of library training on student learning and educational outcomes. Table 1 shows the library training evaluation measures that were most frequently represented in the reviewed literature.

FOUNDATIONS OF INFORMATION LITERACY EVALUATION

During the 1990s, many higher education accrediting agencies adopted IL competencies as important criteria for academic success and achievement [1–4]. At the same time, the Association of College and Research Libraries (ACRL) called for an increased emphasis on IL through its *Information Literacy Competency Standards for Higher Education* [5]. The associated literature described the adoption and practical implications of the ACRL *Standards* and other IL criteria, as well as subsequent IL tests such as Project SAILS or institution-specific IL tests [6–12]. The library literature has also deliberated the educational viability of stand-alone IL tests, IL courses, and IL workshops; curriculum-integrated instruction; the pros and cons of online training; and the like [13–32]. This literature represents the foundations upon which library-based IL education is typically evaluated.

The Association of Research Libraries commissioned a study to document the demonstrated value of academic libraries to their campuses [33, 34]. In the resulting publications, the authors reviewed existing evaluation measures for training and educational programs and discussed the relevance of library instruction for measuring library impact on student learning (2011) [34]. They described options for measuring skills development (2008) via tests, performance, or evaluation rubrics. The research found, though, that too many libraries did not employ systematic approaches to the assessment of student

learning outcomes and IL skills [33]. In fact, many libraries—even those that evaluated their workshops—did not formally evaluate student learning in any way.

This lack of evaluation becomes problematic when libraries must qualify and quantify their impact on educational goals and outcomes. This is an important issue, for without systematic evaluation, libraries do not have adequate information to determine the impact of existing training on students' IL skills and learning needs. Decisions regarding whether educational programs should be continued, expanded, or modified occur in a vacuum. In other cases, libraries may not lack data itself, but may lack a systematic process for using statistically reliable research to make conclusions about the validity of educational programming. This leaves librarians guessing: How do library educational programs impact IL skills development for meaningful student learning outcomes?

TIMING MATTERS: FORMATIVE AND SUMMATIVE EVALUATION

When evaluation activities take place impacts what is measurable, so realistic decisions about formative and summative evaluation are important. *Formative* and *summative* evaluation techniques are applied at different points across the educational cycle. Formative evaluation tracks student progress along the way. Summative evaluation represents a point in time usually either immediately after training or longitudinally. The timing-related pros and cons are discussed here.

Although formative and summative evaluation go hand-in-hand, summative evaluation is more prevalent in the literature. Zhang, Watson, and Banfield analyzed the ERIC literature from 1990 to 2005, finding ten empirical studies that assessed affect, confidence, and attitudes in information seeking [35]. Of these ten, one implemented both pre- (formative) and post-instruction (summative) evaluation. The more current literature includes more articles describing parallel formative and summative evaluation activities like pre- and post-instruction written tests or affective surveys [8, 9, 11, 36–39]. In a well-rounded evaluation system, both formative and summative assessments are essential. Depending too heavily on one or the other may result in skewed perspectives on students and learning, because a complete picture of skills and attitudes is not available.

Formative assessment may be iterative, occurring periodically throughout a training session, course, or workshop series. This allows instruction to be modified on the go. Gilstrap and Dupre described an English composition course in which multiple sections of students (≈ 600) participated in curriculum-integrated IL training [40]. During each of 4 library instruction sessions, students completed "critical incident questionnaires" about their information needs and skills. Reviews of these reports resulted in the development of a list of recurring themes. When preparing for the next session, librarians

reviewed these themes and modified training to address specific problems or areas of confusion.

This iterative approach is advantageous, but not always feasible or practical. For example, it is only practical when there are multiple opportunities to engage with learners across a period of time. However, library instruction may be short of duration (one to three hours) or one time only, limiting opportunities to do formative evaluation and modify training midstream. Time-in-training is another barrier: librarians may sacrifice formative evaluation measures so that more time is available for instruction or hands-on activities.

Summative evaluation is cumulative in nature. It typically occurs immediately after training, which is an obvious window of time to assess students' skills or record their attitudes. Short-term retention should not be used as a substitute for measures of long-term retention, however. Long-term retention is the more accurate indicator of actual learning. Ideally, short-term and longitudinal (weeks, months, semesters later) summative evaluation activities should be strategically implemented to understand more about learners' long-term retention [41]. There is a complex array of benefits and barriers to longitudinal tracking or testing, so finding a balance of measures is important.

Cmor, Chan, and Kong examined literature searching exercises and viewed oral presentations in which students described their search strategies that were recorded immediately post-training and again two weeks after instruction [42]. Content analysis showed that during the 2 weeks immediately following instruction, students reverted to ineffective searching behaviors. Of the 70% of students who showed reasonable competence immediately post-training, "very few" utilized those same effective search techniques just 2 weeks later. The students did not demonstrate learning or retention.

Despite being highly reliable for measuring actual learning, assessing retention longitudinally presents a variety of challenges, which are perhaps evidenced by the small numbers of articles on retention [41, 42]. One drawback is logistical: students have a tendency to scatter after training. Potential sources of information may also be lost over time. Zoellner, Samson, and Hines reported that only 33% of their original 426 communications students participated in summative surveys [38]. Researchers were also unable to track responses across time due to loss of identifying data with the posttest survey. Bronshteyn and Baladad measured learning with individual narrative reflections and anonymous surveys, but neither of these could be correlated because the anonymous surveys were impossible to track [43]. Cmor, Chan, and Kong indicated that 900 students took a skills test immediately post-instruction, but 20 student presentations from 2 weeks later were available for analysis [42]. Figa, Bone, and MacPherson reported that post-course surveys had only an 11% response rate [44]. Blummer also reported problems due to data restrictions and the inability to control for a variety of variables [45].

Table 2
Overview of evaluation measures

Evaluation type		Example
In-process/formative evaluation		
Direct		
Cognitive	Qualitative feedback	■ One-minute question: write what was most confusing about this session.
Behavioral	Observation	■ In lab, observe how many students do or do not utilize specific skills.
	Scored exercises	■ Find a journal article by Smith, M.
Affective*	Pre- and post-questionnaires about confidence levels	■ How do you rate yourself as a searcher?
Indirect		
	Workshop evaluation form	■ How important do you believe this workshop is? How good was the presenter?
	Searching process journals	■ Describe as you go along what you have found for your project.
	Bibliography drafts	■ List key resources (at specified times during project development).
Post/summative assessment†		
Direct		
Cognitive: primarily immediately post-instruction	Test	■ Name two characteristics of a scholarly journal.
Behavioral: primarily long-term/completed project	Critical incident questionnaire	■ Describe the most difficult part of your information search.
	Test/scored exercises	■ Construct a correct Boolean search statement.
	Citation analysis	■ Percent of items from categories of materials.
	Project analysis rubric (presentation or paper)	■ Project uses poor, better, outstanding quality materials.
Affective*	End-of-course or program questionnaire	■ Considering your course tasks, how confident do you feel about your searching skills?
	Focus groups	■ Describe information needs you had and what you did?
	Interviews	■ What were some high or low parts of your information searches this (course, semester, program)?
Indirect		
	Usage (circulation, web hits, downloads)	■ Before and after quantity of use of targeted resources.
	Questions	■ Content analysis of questions received by reference point of service.

* Note about direct-affective measures: Satisfaction or confidence are not direct measures of *skills* (cognitive knowledge or behavioral capability), but are direct measures of participants' *attitudes*.

† Note about short-term versus long-term measures: Tests are usually not feasible except immediately at the end of instruction (within librarian-controlled time). Behavior is somewhat artificial immediately post-instruction, and research on behavior is much more valid when behavior is measured as subsequent real-life use.

These articles provide evidence that without access to individual learners and artifacts, rigorous research methodologies cannot be implemented. Without adequate formative and summative evidence, librarians cannot make reasonable judgments regarding the value and impact of library instruction. Timing matters.

DIRECT AND INDIRECT MEASURES

Evidence of student attitudes and learning can be collected in many ways, including measures of performance (such as tests, portfolios, or products), attitude (surveys), and behavior (library usage). Popular techniques for evaluating the efficacy of library education programs have included in-process and end-product measures, such as citation analysis, narrative reflection, focus groups, and demonstration portfolios and products [43, 45–47]. Evaluation is either *direct* or *indirect*. Direct evaluation documents knowledge, skills, or behavior. Indirect assessment is based on self-reporting of perceived skills, learning, behaviors, or attitudes. The following section provides an overview of the benefits and drawbacks of various commonly cited evaluation approaches. Table 2 provides examples of these evaluation measures.

Measures of performance

Performance measures (direct) are based on actual student work. Most often these include tests or scored exercises grounded in clinical questions or real-life

simulations or displays such as course products, projects, portfolios, research papers, essays, exhibits, case analysis, and so on. Testing is an obvious way to measure skills or knowledge. Portfolio reviews, citation analysis, product reviews, and so on are also useful for understanding what students have achieved or learned over time. Other approaches include reviewing literature searches to assess students' skills in using explode, focus, subheadings, or limits or skills in creating effective search strategies for answering clinical questions [27, 30, 42, 47–55].

One pro of using performance measures is that students already produce course artifacts, including literature searches. Librarians do not necessarily need to implement separate evaluation activities. A negative is that significant coordination with information technology access policies and teaching faculty may be required to gain access to student materials. Another problem is that available course artifacts may not provide the best evidence for accurately assessing information retrieval and IL skills. Plus, intangible variables such as affect, values, perceptions, and beliefs are not represented. These intangibles influence learning.

Knowledge versus skills

In health sciences libraries, a logical approach to measuring skills or knowledge has been through literature search assessment, practical tests or exercises, or written tests. Not all tests are created equal, though. Written tests reflect theoretical knowledge.

Practical exercises or tests demonstrate applied skills [38, 39, 56, 57]. In one experiment, medical students participated in an intervention that included pre- and post-training written tests on literature searching concepts and terminology, types of resources, citation formats, critical analysis, problem solving, and so on. Correlation of written tests with scored MEDLINE searches showed no statistically significant results ($r=+0.17$ not statistically significant) [57]. Theoretical knowledge did not translate into practical skill. Maneuvering through a sophisticated bibliographic database like MEDLINE to find journal literature to address a clinical question requires a unique set of skills [57].

Measures of attitude

Indirect assessment is *affective* or attitudinal, or based on learners' perceptions of their skills and learning, including what people think, feel, or believe about the training experience or about their skills. Sources for these reports are surveys or questionnaires completed by the students themselves, librarians, course instructors, internship supervisors, peers, and so on. Attitudes are quantified in terms of satisfaction, confidence, comfort likes, dislikes, preferences, interests, and so forth. Surveys are often used to evaluate the training itself, the instructors, training resources (blogs, clickers, videos), or the method of instruction (online or face-to-face) [38, 40, 44, 57–60].

Crawford [61] analyzed 215 empirical studies in ERIC from 1970 to 2002, showing that affective surveys were the most frequently used tool for assessing library instruction. In fact, self-reported attitudinal surveys were the most frequently cited method of evaluation in the literature reviewed here [12, 42, 44, 45, 56, 60]. A common "process-implementation" approach is to ask faculty to describe their satisfaction with the quality of student work or products [42]. Although these findings are non-measures in terms of student learning, they are useful for understanding instructor perspectives.

Polkinghorne and Wilton [11] described a political science research skills course in which students' reflections on their information search skills and a variety of course artifacts were reviewed. The analysis of these artifacts provided a window into students' perceptions. One of the more interesting findings was that students focused on end-results and end-products (assignments), rather than on the information retrieval process itself. The authors commented that students, particularly those with weak skills, were not aware of the requirements for improved IL skills [11]. This leads to an important issue in indirect assessment: students may not be excellent self-judges of their own skills [57]. Even when students acknowledged that their searching methods were chaotic and inefficient, they still felt that they had achieved success and were adept researchers [11]. This is indicative of the "satisfied but inept" phenomenon described by Plutchak [62]. Information retrieval may identify only a handful of the available and applicable

resources, and searchers may waste time and go about the process in a haphazard and inefficient manner. Yet, because users find *something* that seems relevant, they are satisfied with their work: thus, the "satisfied but inept" effect.

Related findings were reported in evidence-based information retrieval experiments with first-year and third-year medical students [57, 63]. Triangulation of evaluation results showed students: (1) believed themselves to possess higher IL skills than their test scores indicated, (2) did not know what information resources to use to support their coursework, (3) were not aware of important evidence-based resources, and (4) overestimated their knowledge of those they had used. For example, of the 88% of first-year medical students ($n=128$) who reported being very familiar with MEDLINE, only 12% were aware of Medical Subject Headings (MeSH). In fact, students consistently overestimated their own MEDLINE, information retrieval, and IL skills and knowledge ($r=-0.07$ not statistically significant) [57].

Library and resource usage

Questions about the impact of library training and IL programs are also addressed through library usage data. Two approaches measure library usage: reported usage or actual demonstrated usage. In libraries, resource usage is most often gathered through self-reported surveys [45, 46, 56, 64]. A benefit to this approach is that self-reported library usage behaviors are relevant for understanding attitudes like satisfaction, knowledge of resources, comfort, and so forth. A drawback is that self-reported behavior is based on guessing rather than fact. It may not be a statistically valid marker of actual demonstrated usage [57].

While self-reported library usage behavior does not show actual learning, it may be indicative of the potential for learning. For example, students who are more engaged with library resources have more opportunities to learn to use these resources effectively and efficiently, so perhaps these students will achieve IL higher skills. Unfortunately, the opportunity to learn does not equal actual learning. Another common hypothesis is that students who are more comfortable, satisfied, or familiar with the library are also more likely to use the library. The validity of these claims cannot be determined without real usage data, however; so estimated or reported behaviors are mostly useful for understanding a student's frame of mind.

Demonstrated library usage can be documented through library statistics or data from service areas such as interlibrary loan, circulation, and reserves or from resources including staff, library websites, or bibliographic databases. One benefit of this approach is that it paints a picture of actual resource usage. Another benefit is that usage behavior can be easily tracked longitudinally. One obstacle to measuring user behavior is that it requires tracking of individuals or cohorts. Internal review board requirements may complicate this issue. Plus, many resources are

accessed anonymously, rendering it impossible to track individual users.

When available, library usage data can be correlated with other variables like attitudes, test scores, or student learning outcomes to form a broader picture. For instance, library resource usage by a cohort of first-year medical students ($n=128$) was gathered in 2 ways [27]. Immediately post-training, students were surveyed regarding their anticipated use behavior over the next semester. At the end of the semester, students again described their information use during the previous semester. Actual usage across the semester was also captured using students' individual system accounts. These data were then correlated to training group (online training versus face-to-face training) to determine whether the training method (online or face-to-face) itself impacted expected or demonstrated usage. This was important to librarians who wanted students to "get the message" regarding the importance of particular evidence-based course resources. Data analysis revealed no statistical correlations between training groups. This meant that online instruction and traditional instruction were comparable in terms of their impact on students' anticipated information usage behaviors ($P=0.20$) and actual usage behaviors throughout the semester ($P=0.25$) [27]. These results eased librarians' concerns about online training. Without library usage statistics, this level of comparison would not have been possible.

Implementing the most efficacious evaluation measures and instruments in any given situation can be challenging. Each approach has its own limitations. At the same time, the online learning example above illustrates that not every important research question needs to be about student learning. Evaluation is useful for informing and driving internal decisions and policies. An excellent program of evaluation combines measures from a variety of sources for a variety of research questions. This approach provides a range of artifacts and data from which to draw evidence.

IN-PROCESS AND END-PRODUCT MEASURES

Information retrieval, information literacy, and learning are processes. Capturing the process itself is important. *In-process measures* are useful for identifying exactly where a student's skills are weak or strong. The early foundational literature on information-retrieval skills in biomedical education established the groundwork for understanding the information-retrieval process itself, specifically as it relates to end-user searching in MEDLINE [48–55]. This literature also examined behavioral variables like the extent to which prior MEDLINE searching experience impacted subsequent searching skills and behaviors. For instance, a strong relationship between the level of search experience and frequency of searching was demonstrated [51]. In other words, searching begets searching. Research also showed that searching experience and skill levels correlated. Practice makes

perfect [51]. Additionally, Pao and colleagues found that clinical knowledge did not translate to bibliographic searching skills, evidence that bibliographic searching requires a unique set of skills [50, 51]. Although the information-retrieval process continues to warrant exploration, a strong foundational body of literature addresses many major questions.

In-process evaluation can also be applied to other student artifacts such as presentations, projects, or portfolios. There are three reasonably effective methods for conducting in-process evaluation:

1. An assignment immediately at the end of an instruction session or sequence (summative), particularly those in which students can actually begin their searches (direct), takes minimal course faculty engagement but significant coordination, planning, and grading for librarians. Scored practical exercises are also effective measures of actual skills.
2. A survey administered after a period of time (summative, indirect) requires coordination with course instructors, and a reasonable guarantee that the students are available and willing to complete and return surveys. Self-reported surveys are not strong measures of actual learning.
3. A blog, journal, or portfolio captures elements of the writing process such as initial bibliographies, search journals, notes, and so on (formative and/or summative; direct).

The third method requires significant faculty buy-in and willingness to share student work [9, 11, 12, 40–42, 56, 60, 65]. Process portfolios are also usually only feasible when course projects themselves use a developmental approach that requires evidence such as drafts, journals, or milestone notes or evidence. Unfortunately, many process measures, including those in which course instructors are surveyed about what should be incorporated into an instructional session, are considered "non-measures" for the purpose of evaluating student learning [66, 67].

One of the most commonly used summative end-product approaches to evaluating student learning is the citation analysis method, in which an academic task culminates in a formally cited paper, project, presentation, and so on [10, 12, 41]. Citations are examined for standards of content and quality. This approach has two main strengths. The first is the idea that the paper or product itself is the ultimate goal, similar to the journal article publication of a research scientist. It is a real, authentic summative measure. Secondly, citation analysis is feasible in many academic settings and may require minimal coordination between the content instructor and library instructor.

Sharma's research at the University of Connecticut described a one-credit, stand-alone IL course that used a portfolio, with requirements to incorporate specific searching skills [12]. The primary course goal was to give students hands-on practice and to provide instructors with detailed information for assessment. Portfolios were evaluated with worksheets for each area covered, a rubric system with dimensions (areas to be evaluated) and defined ratings (needs improvement, good, or excellent). Aggregated results showed

Table 3
Summary of evaluation pros and cons

Evaluation tool	Pros	Cons
Performance measures		
Practical exercises	<ul style="list-style-type: none"> ■ Direct measures of skills ■ Can test multiple domains ■ Learner-driven ■ Suited for procedural/technical skills ■ Practical to implement ■ May or may not take class time ■ Easily integrated into lesson ■ Potential for formative and longitudinal trending ■ Evaluator need not be present ■ Direct measures of knowledge 	<ul style="list-style-type: none"> ■ Labor intensive for planning, developing evaluation criteria, grading ■ Samples may not be typical ■ Inter-rater reliability, high subjectivity ■ Disregards affective variables
Tests, exams (locally developed)	<ul style="list-style-type: none"> ■ Can test multiple domains ■ May or may not take class time ■ Potential for formative and longitudinal trending ■ Evaluator need not be present ■ Program and institution-specific ■ Curriculum-related ■ Less costly ■ Local scoring leads to quicker feedback ■ Local control over interpretation and results ■ Reduces faculty time creating instruments ■ Objective results ■ Provides for externality of results ■ Grounded in state, national, professional standards ■ Useful for benchmarking ■ Can test large cohorts 	<ul style="list-style-type: none"> ■ Labor intensive for planning, developing evaluation criteria, grading ■ Inter-rater reliability, high subjectivity ■ Low validity ■ May not provide for externality of results (validity across multiple sites) ■ Not useful for benchmarking or cross-institution comparisons ■ Disregards affective variables ■ Does not measure practical skills ■ Grounded in recall ■ Rewards good test takers ■ Requires significant faculty buy-in ■ Measures superficial knowledge ■ Restricts what can be measured ■ More summative than formative ■ Not discipline- or course-specific, not linked to specific learning outcomes ■ Cost prohibitive ■ Norm-reference data skewed ■ Longer wait time for results ■ Lower validity for group multiple-choice formats ■ Teaching to the exam ■ Requires significant faculty buy-in ■ Existing artifacts may be a poor fit
Standardized IL tests (such as SAILS)	<ul style="list-style-type: none"> ■ Useful for benchmarking ■ Can test large cohorts 	<ul style="list-style-type: none"> ■ Cost prohibitive ■ Norm-reference data skewed ■ Longer wait time for results ■ Lower validity for group multiple-choice formats ■ Teaching to the exam ■ Requires significant faculty buy-in ■ Existing artifacts may be a poor fit
Course products, projects, portfolios, research papers, essays, exhibits, case analysis, citation analysis	<ul style="list-style-type: none"> ■ Learner-driven ■ Encourages active and independent learning 	<ul style="list-style-type: none"> ■ Access to course artifacts may be limited ■ Content may vary widely
Narratives, blogs, journals (for measures of specific content quality)	<ul style="list-style-type: none"> ■ Artifacts already exist ■ Assesses multiple abilities (learners' abilities to generate ideas, weigh arguments, build and support hypotheses, organize information, evaluate information, etc.) ■ Promotes self-reflection and self-assessment ■ Useful for closing student-instructor feedback loop ■ Easily integrated into class lessons ■ Potential for formative and longitudinal trending (with in-process measures) ■ Potential for remediation (with in-process measures) 	<ul style="list-style-type: none"> ■ Labor intensive for developing evaluation criteria and scoring ■ Low inter-rater reliability, high subjectivity ■ Narrow content coverage ■ Does not provide for externality of results ■ Time intensive to convert findings to usable knowledge ■ End-product measures disregard the process
Measures of attitude		
Self-reported surveys	<ul style="list-style-type: none"> ■ Easy to administer ■ Covers multiple topics in brief amount of time ■ Assesses learners' frame of mind ■ Values learners' opinions ■ Results are easily understood by lay persons ■ Useful for triangulation ■ Potential for formative and longitudinal trending 	<ul style="list-style-type: none"> ■ Does not measure skills or knowledge (indirect) ■ Subject to halo effect ■ Survey design is challenging ■ Low response rates ■ Forced-choice questions limit individual responses ■ Bias (volunteer samples) ■ Requires computer analysis for large samples ■ Reflects perceptions (indirect) ■ May not provide for externality of results ■ Logistical issues for implementing pre and post ■ Requires significant faculty buy-in ■ Existing artifacts may be a poor fit ■ Access to artifacts may be limited ■ Content varies widely ■ Time intensive to convert findings to usable knowledge
Narratives, blogs, journals (for student perceptions)	<ul style="list-style-type: none"> ■ Useful evidence of affective variables ■ Useful for formative assessment ■ Minimizes test anxiety ■ Minimizes one-shot measurement problems ■ Learner need not be present during scoring ■ Reflects perceptions and captures process ■ Encourages active and independent learning 	<ul style="list-style-type: none"> ■ Does not measure learning ■ Low reliability ■ Subject to halo effect
Observation	<ul style="list-style-type: none"> ■ Useful evidence of affective variables ■ Simple, efficient ■ Covers multiple topics in brief amount of time ■ Assesses learners' frame of mind ■ Values learners' opinions 	<ul style="list-style-type: none"> ■ Does not measure learning ■ Low reliability ■ Subject to halo effect
Interviews, focus groups	<ul style="list-style-type: none"> ■ Useful evidence of affective variables ■ Gathers ideas, details, insights ■ Inexpensive ■ Helps with interpretation of results from other measures ■ Unanticipated issues can be explored 	<ul style="list-style-type: none"> ■ Requires trained moderator ■ Suitable for small samples ■ Groups are difficult to assemble ■ Subject to halo effect ■ Data are complex to analyze ■ Not suited for making generalizations

Table 3
continued

Evaluation tool	Pros	Cons
Behavior: library and resource usage		
Self-reported usage	<ul style="list-style-type: none"> ■ Potential for formative and longitudinal trending ■ Assesses learners' frame-of-mind ■ Values learners' opinions 	<ul style="list-style-type: none"> ■ Does not measure skills or knowledge (indirect) ■ Subject to halo effect
Demonstrated usage (data collected)	<ul style="list-style-type: none"> ■ Easy to administer ■ Easily collected and analyzed ■ Potential for formative and longitudinal trending 	<ul style="list-style-type: none"> ■ Does not measure skills or knowledge (indirect) ■ Existing data may be poor fit ■ Requires ability to track individuals or cohort ■ Requires computer analysis for large samples ■ May not provide for externality of results

that a particularly difficult step in the process was in defining a topic in a way that was conducive to developing an effective search strategy [12]. This phenomenon is routinely observed among librarians who spend considerable time teaching students about appropriate "research questions" and "searchable questions."

There are two related drawbacks to the end-product evaluation approach. One is that a final product does not provide information about the information retrieval process itself. Library instruction aims at producing searchers who are both effective and efficient. A student who is a poor searcher but a good evaluator, for example, could produce a high-quality citation list only after wading through hundreds of citations. An effective searcher would have achieved the same final results, but with much greater precision and, hence, economy of effort. Citation analysis loses this important information.

Another drawback to end-product citation analysis is the loss of those citations and resources that are used but may not appear on a final bibliography. An anthropological observation study by Edge showed that astronomers read, talked about, and were influenced by far more than just those citations that appeared in their formal papers [68]. Students are likewise influenced by items that they retrieve and read. Another drawback is that citation analysis does not describe where the search process has failed: poor use of Boolean logic, ineffective application of limit options, poor article evaluation skills, and so on. Citation analysis overlooks the searching process and does not measure actual searching behavior or information retrieval skills. Table 3 summarizes the pros and cons of commonly used evaluation measures.

ALL MEASURES ARE NOT CREATED EQUAL

It is perhaps most important to distinguish between "what one wants to measure" versus "what really is being measured." While there are many important distinctions between direct testing and affective and behavioral approaches to evaluation, the complex relationships between measures remain unclear. A report in the field of engineering education, for instance, concluded that self-reports demonstrated a "very good relationship" to objective measures [69]. Weisskirch and Silveria reported that higher student

confidence in the process (e.g., the ability to evaluate the quality of sources) correlated positively with higher grades on a final project [60]. Kisby reported that college students who rated their "online skills" as high also reported greater satisfaction with library services [64]. Lower self-assessed skills were associated with lower satisfaction and nonuse of services. In contrast to these findings, Schilling and Applegate found that attitudes neither predicted nor equaled skill [57]. Students' ($n=128$) MEDLINE and information retrieval skills (graded search exercise) did not statistically significantly correlate to any of the following:

1. student self-reported satisfaction with their searches ($r=+0.12$ not statistically significant)
2. confidence/comfort in their searches ($r=+0.14$ not statistically significant)
3. frustration with the information retrieval process ($r=-0.09$ not statistically significant)

Interestingly, students did know when they had learned something. Those who reported learning more during the training did achieve higher MEDLINE searching skills ($r=+0.41$ statistically significant) [57]. Contradictory research results leave unanswered questions about the efficacy of affective measures.

Affective measures are important because user satisfaction and self-confidence play into information behavior and skills development. At the same time, the common affective measures, like questionnaires and self-reporting skills surveys, cannot substitute as reliable measures of skills and knowledge. The challenge becomes to implement a balanced approach. This requires that librarians distinguish between and systematically apply appropriate non-measures of learning (faculty observations), indirect measures of learning (self-assessment of skills), and direct measures of learning (tests).

Another issue that presented in the literature review had to do with the nature of statistical analysis. Most of the literature exclusively reports simple descriptive statistics. These measure one variable at a time, independent of one another. Descriptive statistics are not particularly effective for revealing complex relationships among multiple variables. Moving beyond simple descriptive analysis to the triangulation of data from multiple sources can show the extent to which different evaluation measures are interdependent, which evaluation measures have cause-and-effect relationships, and, most

importantly for learning outcomes, what combinations of measures most accurately represent skills, knowledge, and learning. Librarians need not become statisticians themselves but can periodically consult with statisticians to identify options for uncovering deeper relationships within the myriad data that they may already have on hand. Tables 4 and 5 show statistical results from a previous study that illustrate the efficacy of tests, attitudes, reported behavior, and other variables for assessing learning [57]. These findings revealed relationships among multiple variables for understanding learning itself and attitudes about learning.

These results and the results of other studies reviewed above indicated that:

- Practical exercises were the most efficacious way to document actual applied, practical skills.
- Written tests were not the equivalent of a practical searching exercise in terms of measuring applied, practical skills.
- Students were poor self-judges of their own skill levels. In fact, their attitudes evidenced the “satisfied but inept” phenomenon [62].
- Students were excellent self-judges of their attitudes, feelings, beliefs, and perceptions. Certain attitudes went hand-in-hand, like confidence and satisfaction or frustration and dissatisfaction.
- Attitudes, feelings, beliefs, and perceptions were not indicative of actual knowledge and learning.

This level of statistical analysis provided evidence upon which to make conclusions about the impact of library training on learners’ information usage, attitudes, skills, and behavior and, ultimately, on student learning and achievement of educational outcomes.

DISCUSSION AND CONCLUSION

This paper provides an overview of current approaches to evaluating library instruction and students’ achievement of educational outcomes, describing the most commonly used evaluation measures among libraries and presenting pros and cons or benefits and barriers of each. This paper also illustrates the significant disparities between measures of skills,

knowledge, attitudes, and information usage behaviors. While affective measures are more common, they are not likely to provide meaningful evidence in terms of students’ skills, course grades, or learning outcomes. Yet, attitudes do remain important in the larger context. Librarians want students to feel comfortable and confident in the library and to learn to identify their own limitations and learning needs. Academic assessment protocols strongly urge the incorporation of direct testing measures, such as objective tests or expert reviews of performances (searches), artifacts, or portfolios [70]. Although every evaluation measure has value, retention of learning is evidenced most efficaciously through demonstrated skills that are assessed longitudinally. When direct measures, such as tests, are expensive, cumbersome to implement, or difficult to analyze, indirect measures are often accepted as viable alternatives. Evaluators must avoid confusing “material” satisfaction (whether the learner *demonstrated* effective skills) versus “emotional” satisfaction (how the learner *felt* about his or her skills). One must consider also the issue of *validity* of measures. This requires a deliberate and strategic approach to evaluation, asking: What questions do we have? What evidence do we need? What can we realistically evaluate? What data can we realistically collect? Through what methods and tools or instruments can these data be realistically collected?

Libraries continue to invest significant time and attention to educational programs and know that systematic and sequential information literacy education is important to problem solving, critical thinking, and lifelong learning. Likewise, librarians must equally evaluate their training and educational program outcomes systematically and sequentially with rigorous research methods and measures. The goal of effective evaluation remains to implement what is known about the efficacy of evaluation methods, measures, and tools; weigh the pros and cons, benefits and barriers of each; and acknowledge the required compromises for feasibility, practicality, and affordability. The efficacy of the evaluation measures employed will continue to have a major impact on what questions are answered and unan-

Table 4
Efficacy of evaluation measures*

Evaluation measure and method	Did measure	Did not measure
1. Applied, demonstrated skills, scored practical searching exercise, conducted during the week after training	Practical, applied skills	Long-term retention; written, theoretical knowledge, attitudes, preferences, comfort, perceptions, beliefs; behavior, library/information usage
2. Theoretical knowledge, pre- and post-training written tests	Written, theoretical knowledge; learning from pre- to post-instruction	Practical, applied skills; long-term retention; attitudes, preferences, comfort, perceptions, beliefs; behavior, library/information usage
3. Self-reported information retrieval skills	Attitudes, preferences, comfort, perceptions, beliefs	Practical, applied skills
4. Confidence		Long-term retention
5. Satisfaction		Written, theoretical knowledge
6. Frustration		Behavior, library/information usage
7. Workshop evaluation; pre- and post-instruction surveys		

* Based on data from Schilling and Applegate [57].

Table 5
Results of statistical correlations*

Evaluation measure	Results
1 and 2	$r=+0.17$ (not statistically significant (n.s.s.))
1 and 3	$r=-0.07$ (n.s.s.)
1 and 4	$r=+0.12$ (n.s.s.)
1 and 5	$r=+0.06$ (n.s.s.)
4 and 5	$r=+0.23$ (statistically significant (s.s.))
5 and 6	$r=+0.44$ (s.s.)
3 and 7	$r=+0.23$ (s.s.)
5 and 7	$r=+0.41$ (s.s.)

* Based on data from Schilling and Applegate [57].

swered, and on what evidence is available to quantify the educational value of library instruction.

REFERENCES

1. New England Association of Schools and Colleges. Standards for accreditation. Winchester, MA: New England Association of Schools and Colleges, Commission on Institutes of Higher Education; 1992.
2. Southern Association of Colleges and Schools. Criteria for accreditation: commission on colleges. Atlanta, GA: Southern Association of Colleges and Schools; 1992.
3. Association of American Medical Colleges. Medical School Objectives Project (MSOP): report IV: contemporary issues in medicine: basic science and clinical research [Internet]. The Association; 2001 Aug. p. 24. [cited 29 Apr 2012]. <<https://www.aamc.org/initiatives/msop>>.
4. Association of American Medical Colleges. Physicians for the twenty-first century: report of the panel on the General Professional Education of the Physician and college preparation for medicine (GPEP). *J Med Educ*. 1984;59(11, 2): 1-208.
5. American Library Association, Association of College and Research Libraries. Information literacy competency standards for higher education: standards, performance indicators, and outcomes [Internet]. The Association; 2000. 23 paragraphs. [cited 29 Apr 2012]. <<http://www.ala.org/acrl/standards/informationliteracycompetency>>.
6. Kent State University. Project SAILS (Standardized Assessment of Information Literacy Skills) [Internet]. The University; (2000-2012) [cited 29 Apr 2012]. <<https://www.projectsails.org>>.
7. Lym B, Grossman H, Yannotta L, Talih M. Assessing the assessment: how institutions administered, interpreted, and used SAILS. *Ref Serv Rev*. 2009;38(1):168-86.
8. Cameron L, Wise SL, Lottridge SM. The development and validation of the information literacy test. *Coll Res Lib*. 2007 May;68(3):229-36.
9. Burkhardt JM. Assessing library skills: a first step to information literacy. *Portal: Lib Acad*. 2007;7(1):25-49.
10. Scharf D, Elliot N, Huey HA, Briller V, Joshi K. Direct assessment of information literacy using writing portfolios. *J Acad Librariansh*. 2007 Jul;33(4):462-77.
11. Polkinghorne S, Wilton S. Research is a verb: exploring a new information literacy-embedded undergraduate research methods course. *Can J Inf Lib Sci*. 2010 Dec;34(4):457-73.
12. Sharma S. Perspectives on: from chaos to clarity: using the research portfolio to teach and assess information literacy skills. *J Acad Librariansh*. 2007 Jan;33(1):127-35.
13. Rui W. The lasting impact of a library credit course. *Portal: Lib Acad*. 2006;6(1):79-92.
14. Breivic PS. Putting libraries back in the information society. *Am Lib*. 1985 Nov;16:723.
15. Breivic PS. Making the most of libraries in the search for academic excellence. *Change*. 1987 Jul-Aug;19:46.
16. Breivic PS, Gee EG. Information literacy: revolution in the library. New York, NY: Macmillan; 1989.
17. Mellon CA. Process not product in course-integrated instruction: a generic model of library research. *Coll Res Lib*. 1984 Nov;45(6):471-8.
18. Mellon CA, ed. Bibliographic instruction: the second generation. Englewood, CO: Libraries Unlimited; 1987.
19. Nowicki S. Information literacy and critical thinking in the electronic environment. *J Instr Deliv Syst*. 1999 Winter;13(1):25-8.
20. Foust JE, Tannery NH, Detlefsen EG. Implementation of a web-based tutorial. *Bull Med Lib Assoc*. 1999 Oct; 87(4):477-9.
21. Kelley KB, Orr GL, Houck J, Schweber C. Library instruction for the next millennium: two web-based courses to teach distant students information literacy. *J Lib Adm*. 2001;32(1-2):281-94.
22. Lindsay EB. Distance teaching: comparing two online information literacy courses. *J Acad Librariansh*. 2004; 30:482-7.
23. Maughan PD. Assessing information literacy among undergraduates: a discussion of the literature and the University of California-Berkeley assessment experience. *Coll Res Lib*. 2001;62(1):71-85.
24. Onwuegbuzie AJ, Jiao QG. The relationship between library anxiety and learning styles among graduate students: implications for library instruction. *Lib Inf Sci Res*. 1998;20(3):235-49.
25. Ren WH. Library instruction and college student self-efficacy in electronic information searching. *J Acad Librariansh*. 2000;26(5):323-8.
26. Stamatoplos A, Mackoy R. Effects of library instruction on university students' satisfaction with the library: a longitudinal study. *Coll Res Lib*. 1998;59(4):323-34.
27. Schilling K. Information retrieval skills development in electronic learning environments: the impact of training method on students' learning outcomes, information usage

- patterns, and attitudes. *IADIS Int J WWW/Internet*. 2006; 4(1):27–42.
28. Green BF, Lin W, Nollan R. Web tutorials: bibliographic instruction in a new medium. *Med Ref Serv Q*. 2006; 25(1):83–91.
 29. Lindsay EB, Cummings L, Johnson CM, Scales BJ. If you build it, will they learn? assessing online information literacy tutorials. *Coll Res Lib*. 2006;67(5):429–45.
 30. Shute SJ, Smith PJ. Knowledge-based search tactics. *Inf Process Manag*. 1993;29(1):29–45.
 31. Lindsay EB. Distance teaching: comparing two online information literacy courses. *J Acad Librariansh*. 2004; 30:482–7.
 32. Fafeita J. The current status of teaching and fostering information literacy in Tafe. *Aust Acad Res Lib*. 2006; 37(2):136–61.
 33. Oakleaf M. Dangers and opportunities: a conceptual map of information literacy assessment approaches. *Portal: Lib Acad*. 2008;8(3):233–53.
 34. Oakleaf M. Are they learning? are we? learning outcomes and the academic library. *Lib Q*. 2011 Jan; 2011(81):61–82.
 35. Zhang L, Watson EM, Banfield L. The efficacy of computer-assisted instruction versus face-to-face instruction in academic libraries: a systematic review. *J Acad Lib*. 2007 Jul;33(4):478–84.
 36. Swoger BJM. Closing the assessment loop using pre- and post-assessment. *Ref Serv Rev*. 2011 May;39(2):244–59.
 37. Hufford JR. What are they learning? pre- and post-assessment surveys for LIBR 1100, Introduction to Library Research. *Coll Res Lib*. 2010 Mar;71(2):139–58.
 38. Zoellner K, Samson S, Hines S. Continuing assessment of library instruction to undergraduates: a general education course survey research project. *Coll Res Lib*. 2008 Jul; 69(4):370–83.
 39. Staley SM, Branch NA, Hewitt TL. Standardised library instruction assessment: an institution-specific approach. *Info Res [Internet]*. 2010 Sep;15(3) [cited 12 Jun 2012]. <<http://www.informationr.net/ir/15-3/paper436.html>>.
 40. Gilstrap DL, Dupre J. Assessing learning, critical reflection, and quality educational outcomes: the critical incident questionnaire. *Coll Res Lib*. 2008 Sep;69(6):407–26.
 41. Barratt CC, Nielsen K, Desmet C, Balthazor R. Collaboration is key: librarians and composition instructors analyze student research and writing. *Portal: Lib Acad*. 2009;9(1):37–46.
 42. Cmor D, Chan A, Kong T. Course-integrated learning outcomes for library database searching: three assessment points on the path to evidence. *Evid Based Libr Inf Pract*. 2010;5(1):64–81.
 43. Bronshteyn K, Baladad R. Librarians as writing instructors: using paraphrasing exercises to teach beginning information literacy students. *J Acad Librariansh*. 2006;32(5):533–6.
 44. Figa E, Bone T, MacPherson JR. Faculty-librarian collaboration for library services in the online classroom: student evaluation results and recommended practices for implementation. *J Lib Info Serv Dist Learn*. 2009;3(2):67–102.
 45. Blummer B. Assessing patron learning from an online library tutorial. *Community Jr Coll Lib*. 2007;14(2):121–38.
 46. Hernon P. What really are student learning outcomes? pushing the edge: explore, engage, extend. *Proceedings of the Fourteenth National Conference of the Association of College and Research Libraries*; Seattle, WA; 2009. p. 28–35.
 47. Canick S. Legal research [instruction] assessment. *Legal Ref Serv Q*. 2009;28:201–17.
 48. Bates MJ. Search and idea tactics. In: White HD, Bates MJ, Wilson P, eds. *For information specialists: interpretations of reference and bibliographic work*. Norwood, NJ: Ablex; 1992. p. 183–200.
 49. Fidel R. Moves in online searching. *Online Rev*. 1985;9(1):61–74.
 50. Pao ML, Grefsheim SF, Barclay ML, Wooliscroft JO, McQuillan M, Shipman BL. Factors affecting students' use of MEDLINE. *Comput Biomed Res*. 1993;26:541–55.
 51. Pao ML, Grefsheim SF, Barclay ML, Wooliscroft JO, Shipman BL, McQuillan M. Effect of search experience on sustained MEDLINE usage by students. *Acad Med*. 1994;69(11):914–9.
 52. Wildemuth BM, Moore ME. End user searching of Medline. Final report. Washington, DC: Council on Library Resources; 1993 Aug.
 53. Wildemuth BM, Moore ME. End-user search behaviors and their relationship to search effectiveness. *Bull Med Lib Assoc*. 1995 Jul;83(3):294–304.
 54. Rosenberg WMC, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. *J Royal Coll Physicians*. 1998;32(6):557–63.
 55. Burrows SC, Tylman V. Evaluating medical student searches of MEDLINE for evidence-based information: process and application of results. *Bull Med Lib Assoc*. 1999 Oct;87(4):471–6.
 56. Sobel K, Wolf K. Updating your tool belt: redesigning assessments of learning in the library. *Ref User Serv Q*. 2011;50(3):245–58.
 57. Schilling K, Applegate RA. Evaluating library instruction: measures for assessing educational quality and impact. *Proceedings of the Thirteenth National Conference of the Association of College and Research Libraries*; Baltimore, MD; 2007. p. 206–14.
 58. Appelt KM, Pendell K. Assess and infest: faculty feedback on library tutorials. *Coll Res Lib*. 2010 May;71(3):245–53.
 59. Ariew S, Lener E. Evaluating instruction: developing a program that supports the teaching librarian. *Res Strat*. 2007;20:506–15.
 60. Weisskirch RS, Silveria JB. The effectiveness of project-specific information competence instruction. *Res Strateg*. 2007;20:370–8.
 61. Crawford J. The use of electronic information services and information literacy: a Glasgow Caledonian University study. *J Librariansh Inf Sci*. 2006;38(1):33–44.
 62. Plutchak TS. On the satisfied and inept end user. *Med Ref Serv Q*. 1989 Spring;8(1):45–8.
 63. Schilling K, Wiecha J, Polineni D, Khalil S. An interactive web-based curriculum on evidence-based medicine: design and effectiveness. *Fam Med*. 2006 Feb;38(2):126–32.
 64. Kisby CM. Self-assessed learning and user-satisfaction in regional campus libraries. *J Acad Librariansh*. 2011;37(6):523–31.
 65. Xin W, Kwangsu C. Computational linguistic assessment of genre differences focusing on text cohesive devices of student writing: implications for library instruction. *J Acad Librarianshp*. 2010 Nov;36(6):501–10.
 66. Palomba CA, Banta TW. *Assessment essentials: planning, implementing, and improving assessment in higher education*. San Francisco, CA: Jossey-Bass; 1999.
 67. Applegate R. Models of user satisfaction: understanding false positives. *RQ*. 1993;32(4):525–39.
 68. Edge D. Quantitative measures of communication in science: a critical review. *History Sci*. 1979;17(35):102–34.
 69. Lattuca L, Terenzini PT, Volkwein JF. Engineering change: a study of the effect of Ec2000: executive summary [Internet]. ABET; 2006 [cited 29 Apr 2012]. <http://www.abet.org/uploadedFiles/Publications/Special_Reports/Engineering_Change-executive-summary.pdf>.
 70. Suskie L. *Assessing student learning: a common sense guide*. San Francisco, CA: Jossey-Bass; 2009.

AUTHORS' AFFILIATIONS

Katherine Schilling, MLS, EdD, katschil@iupui.edu, Associate Professor of Library and Information Science; **Rachel Applegate, MLS, PhD**, rapplega@iupui.edu, Associate Professor of Library and

Information Science; School of Library and Information Science, Indiana University, 755 West Michigan Street, UL3100A, Indianapolis, IN 46202

Received February 2012; accepted May 2012